Online Bookstore Analysis – Amazon Bestsellers (2009-2019)

Overview:

A simple and effective bookstore analysis to help users buy the top bestsellers (2009-2019) on Amazon based on user ratings and reviews.

Data:

I used a dataset from Kaggle: [Amazon Bestsellers - Raw Data](#) (CSV file).

Data Cleaning:

I performed data cleaning on the above raw data using MS Excel. My clean-up tasks included the following steps:
- o Performed a basic spellcheck to remove unwanted special characters, correct misspelled words, etc.
- o Edited three of the column header names: ("Name" to "book_name", "user rating" to "user_rating", "price" to "price_in_$") to ensure compatibility with the database schema.
- o Removed duplicate entries from name column - 200 rows of duplicate entries were removed.
- o Researched and updated correct names for a few incorrect book and author names.

(Refer [File 1](#))

After the clean-up, I created a table schema in a PostgreSQL database and imported the cleaned-up CSV file to the database via pgAdmin 4 using the "Import/Export Data" tool. After the import, for a further clean-up, I checked for null values in each column in the table using SQL queries.

Univariate Exploratory Data Analysis:

I performed an exploratory data analysis in SQL to answer the following questions:
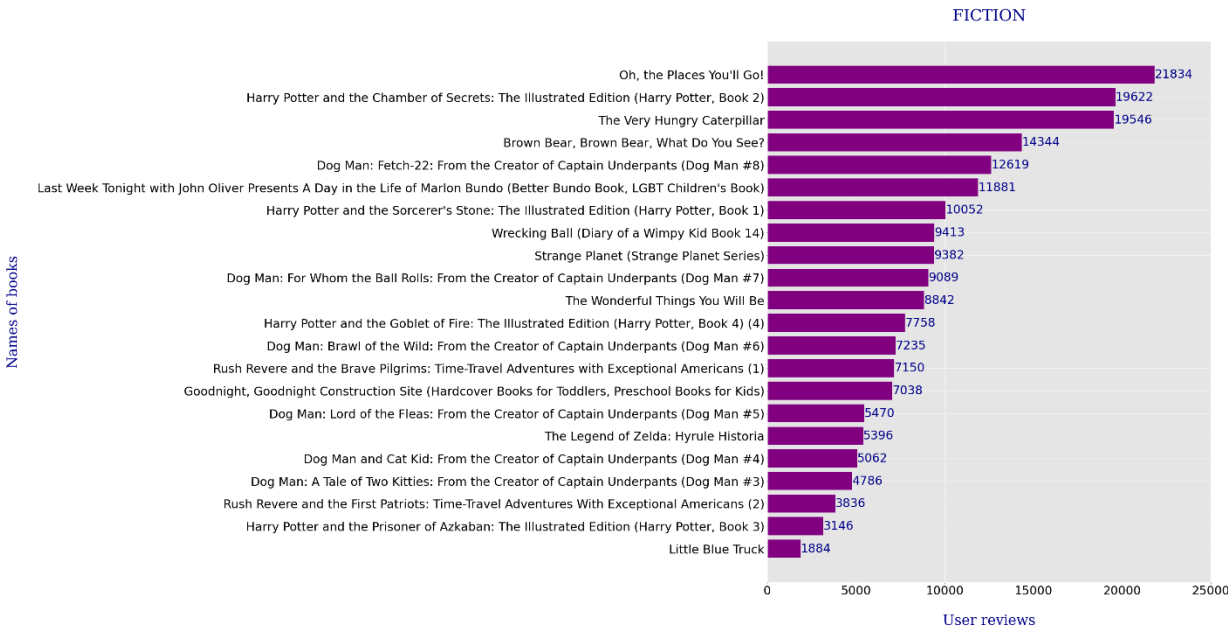
1) Which books have the highest user rating in both fiction and non-fiction genre?
2) From the results of the 1st question, what are the most popular books based on user reviews (in descending order) in both fiction and non-fiction genre?

(Refer [File 2](#))

Data Visualization (Results):

I used Python for data visualization (Refer [File 3](#) & [File 4](#)). Here are the results:

FICTION

Book	User reviews
Oh, the Places You'll Go!	21834
Harry Potter and the Chamber of Secrets: The Illustrated Edition (Harry Potter, Book 2)	19622
The Very Hungry Caterpillar	19546
Brown Bear, Brown Bear, What Do You See?	14344
Dog Man: Fetch-22: From the Creator of Captain Underpants (Dog Man #8)	12619
Last Week Tonight with John Oliver Presents A Day in the Life of Marlon Bundo (Better Bundo Book, LGBT Children's Book)	11881
Harry Potter and the Sorcerer's Stone: The Illustrated Edition (Harry Potter, Book 1)	10052
Wrecking Ball (Diary of a Wimpy Kid Book 14)	9413
Strange Planet (Strange Planet Series)	9382
Dog Man: For Whom the Ball Rolls: From the Creator of Captain Underpants (Dog Man #7)	9089
The Wonderful Things You Will Be	8842
Harry Potter and the Goblet of Fire: The Illustrated Edition (Harry Potter, Book 4) (4)	7758
Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man #6)	7235
Rush Revere and the Brave Pilgrims: Time-Travel Adventures with Exceptional Americans (1)	7150
Goodnight, Goodnight Construction Site (Hardcover Books for Toddlers, Preschool Books for Kids)	7038
Dog Man: Lord of the Fleas: From the Creator of Captain Underpants (Dog Man #5)	5470
The Legend of Zelda: Hyrule Historia	5396
Dog Man and Cat Kid: From the Creator of Captain Underpants (Dog Man #4)	5062
Dog Man: A Tale of Two Kitties: From the Creator of Captain Underpants (Dog Man #3)	4786
Rush Revere and the First Patriots: Time-Travel Adventures With Exceptional Americans (2)	3836
Harry Potter and the Prisoner of Azkaban: The Illustrated Edition (Harry Potter, Book 3)	3146
Little Blue Truck	1884

Names of books / User reviews

More details on the above FICTION bestsellers

Name of the book	Author	Price($)	Bestselling Year
Oh, the Places You'll Go!	Dr. Seuss	8	2012
Harry Potter and the Chamber of Secrets: The Illustrated Edition (Harry Potter, Book 2)	J.K. Rowling	30	2016
The Very Hungry Caterpillar	Eric Carle	5	2013
Brown Bear, Brown Bear, What Do You See?	Bill Martin Jr.	5	2017
Dog Man: Fetch-22: From the Creator of Captain Underpants (Dog Man #8)	Dav Pilkey	8	2019
Last Week Tonight with John Oliver Presents A Day in the Life of Marlon Bundo (Better Bundo Book, LGBT Children's Book)	Jill Twiss	13	2018
Harry Potter and the Sorcerer's Stone: The Illustrated Edition (Harry Potter, Book 1)	J.K. Rowling	22	2016
Wrecking Ball (Diary of a Wimpy Kid Book 14)	Jeff Kinney	8	2019
Strange Planet (Strange Planet Series)	Nathan W. Pyle	6	2019
Dog Man: For Whom the Ball Rolls: From the Creator of Captain Underpants (Dog Man #7)	Dav Pilkey	8	2019
The Wonderful Things You Will Be	Emily Winfield Martin	10	2016
Harry Potter and the Goblet of Fire: The Illustrated Edition (Harry Potter, Book 4) (4)	J. K. Rowling	18	2019
Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man #6)	Dav Pilkey	4	2018
Rush Revere and the Brave Pilgrims: Time-Travel Adventures with Exceptional Americans (1)	Rush Limbaugh	12	2013
Goodnight, Goodnight Construction Site (Hardcover Books for Toddlers, Preschool Books for Kids)	Sherri Duskey Rinker	7	2012
Dog Man: Lord of the Fleas: From the Creator of Captain Underpants (Dog Man #5)	Dav Pilkey	6	2018
The Legend of Zelda: Hyrule Historia	Patrick Thorpe	20	2013
Dog Man and Cat Kid: From the Creator of Captain Underpants (Dog Man #4)	Dav Pilkey	6	2018
Dog Man: A Tale of Two Kitties: From the Creator of Captain Underpants (Dog Man #3)	Dav Pilkey	8	2017
Rush Revere and the First Patriots: Time-Travel Adventures With Exceptional Americans (2)	Rush Limbaugh	12	2014
Harry Potter and the Prisoner of Azkaban: The Illustrated Edition (Harry Potter, Book 3)	J.K. Rowling	30	2017
Little Blue Truck	Alice Schertle	0	2014

NON FICTION

Top rated most popular bestsellers on Amazon (2009-2019) — NON FICTION

	User reviews
Jesus Calling: Enjoying Peace in His Presence (with Scripture References)	19576
The Magnolia Story	7861
Unfreedom of the Press	5956
Hamilton: The Revolution	5867
Obama: An Intimate Portrait	3192
Humans of New York : Stories	2812

(Names of books vs User reviews)
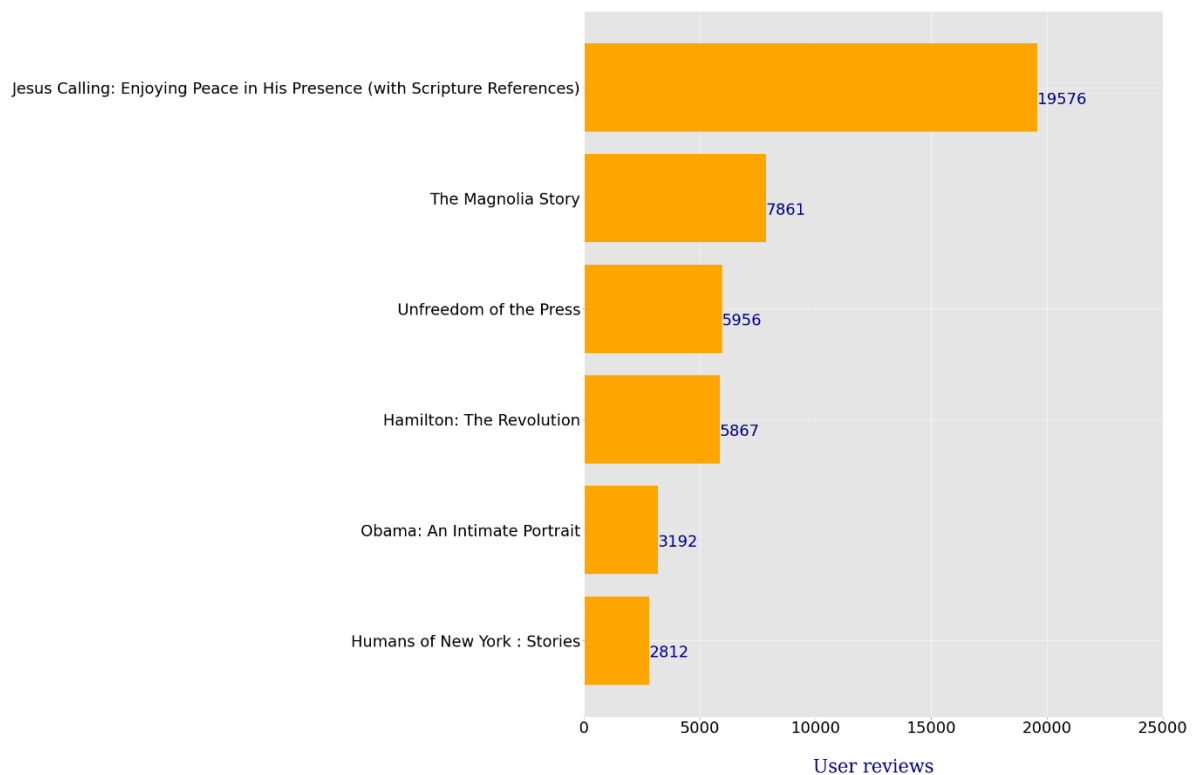
More details on the above NON FICTION bestsellers

Name of the book	Author	Price($)	Bestselling Year
Jesus Calling: Enjoying Peace in His Presence (with Scripture References)	Sarah Young	8	2011
The Magnolia Story	Chip Gaines	5	2016
Unfreedom of the Press	Mark R. Levin	11	2019
Hamilton: The Revolution	Lin-Manuel Miranda	54	2016
Obama: An Intimate Portrait	Pete Souza	22	2017
Humans of New York : Stories	Brandon Stanton	17	2015

Conclusion:

In conclusion, there were a total of 22 bestsellers (2009-2019) with the highest user rating of 4.9 in the **fiction** genre. Among those 22, the most popular book based on user reviews was "Oh, the Places You'll Go!" written by Dr. Seuss. It was an Amazon bestseller in the year 2012 priced at $8.

By contrast, there were only 6 bestsellers (2009-2019) with the highest user rating of 4.9 in the **non-fiction** genre. Among those 6, the most popular book based on user reviews was "Jesus Calling: Enjoying Peace in His Presence (with Scripture References)" written by Sarah Young. It was an Amazon bestseller in the year 2011priced at $8.

Software:

- o MS Excel (version 2303)
- o Database: PostgreSQL 15 (language: SQL, GUI: pgAdmin 4)
- o IDE: Jupyter Notebook (language: Python 3.6.5)
- o Python libraries: psycopg2, matplotlib's pyplot, matplotlib's PdfPages

Files attached:

1. Amazon Bestsellers – Post Cleanup.csv
2. Data Analysis.sql
3. Data Visualization.ipynb
4. Results.pdf